A Generalization of Repetition Threshold

Lucian Ilie*
Department of Computer Science
University of Western Ontario
London, ON, N6A 5B7
CANADA
ilie@csd.uwo.ca

Jeffrey Shallit†
School of Computer Science
University of Waterloo
Waterloo, ON, N2L 3G1
CANADA
shallit@graceland.uwaterloo.ca

February 1, 2008

Abstract

Brandenburg and (implicitly) Dejean introduced the concept of *repetition threshold*: the smallest real number α such that there exists an infinite word over a k-letter alphabet that avoids β-powers for all $\beta > \alpha$. We generalize this concept to include the lengths of the avoided words. We give some conjectures supported by numerical evidence and prove one of these conjectures.

1 Introduction

In this paper we consider some variations on well-known theorems about avoiding repetitions in words.

A *square* is a repetition of the form xx, where x is a nonempty word; an example in English is `hotshots`. It is easy to see that every word of length ≥ 4 over an alphabet of two letters must contain a square, so squares cannot be avoided in infinite binary words. However, Thue showed [14, 15, 2] that there exist infinite words over a three-letter alphabet that avoid squares.

Instead of avoiding *all* squares, one interesting variation is to avoid *all sufficiently large* squares. Entringer, Jackson, and Schatz [7] showed that there exist infinite binary words avoiding all squares xx with $|x| \geq 3$. Furthermore, they proved that every binary word of length ≥ 18 contains a factor of the form xx with $|x| \geq 2$, so the bound 3 is best possible. For some other papers about avoiding sufficiently large squares, see [6, 11, 8, 12, 13].

Another interesting variation is to consider avoiding fractional powers. For $\alpha \geq 1$ a rational number, we say that y is an α-power if we can write $y = x^n x'$ with x' a prefix of x and $|y| = \alpha|x|$. For example, the word `alfalfa` is a 7/3-power and the word `tormentor` is a $\frac{3}{2}$-power. For real $\alpha > 1$, we say a word *avoids α-powers* if it contains no factor that is a α'-power for any rational $\alpha' \geq \alpha$. Brandenburg [3] and (implicitly) Dejean [5] considered the problem of determining the *repetition threshold*; that is, the least exponent $\alpha = \alpha(k)$ such that there exist infinite words over an alphabet of size k that avoid $(\alpha + \epsilon)$-powers for all $\epsilon > 0$. Dejean proved that $\alpha(3) = 7/4$. She also

*Research supported in part by NSERC grant R3143A01.
†Research supported in part by NSERC.

conjectured that $\alpha(4) = 7/5$ and $\alpha(k) = k/(k-1)$ for $k \geq 5$. In its full generality, this conjecture is still open, although Pansiot [10] proved that $\alpha(4) = 7/5$ and Moulin-Ollagnier [9] proved that Dejean's conjecture holds for $5 \leq k \leq 11$. For more information, see [4].

In this paper we consider combining these two variations. We generalize the repetition threshold of Dejean to handle avoidance of all sufficiently large fractional powers. (Pansiot also suggested looking at this generalization at the end of his paper [10], but to the best of our knowledge no one else has pursued this question.) We give a large number of conjectures, supported by numerical evidence, about generalized repetition threshold, and prove one of them.

2 Definitions

Let $\alpha > 1$ be a rational number, and let $\ell \geq 1$ be an integer. A word w is a *repetition of order* α *and length* ℓ if we can write it as $w = x^n x'$ where x' is a prefix of x, $|x| = l$, and $|w| = \alpha|x|$. For brevity, we also call w a (α, ℓ)-*repetition*. Notice that an α-power is an (α, ℓ)-repetition for some ℓ. We say a word is (α, ℓ)-*free* if it contains no factor that is a (α', ℓ')-repetition for $\alpha' \geq \alpha$ and $\ell' \geq \ell$. We say a word is (α^+, ℓ)-*free* if it is (α', ℓ)-free for all $\alpha' > \alpha$.

For integers $k \geq 2$ and $\ell \geq 1$, we define the *generalized repetition threshold* $R(k, \ell)$ as the real number α such that either

(a) over a k-letter alphabet there exists an (α^+, ℓ)-free infinite word, but all (α, ℓ)-free words are finite; or

(b) over a k-letter alphabet there exists a (α, ℓ)-free infinite word, but for all $\epsilon > 0$, all $(\alpha - \epsilon, \ell)$-free words are finite.

Notice that $R(k, 1)$ is essentially the repetition threshold of Dejean and Brandenburg.

Theorem 1 *The generalized repetition threshold $R(k, \ell)$ exists and is finite for all integers $k \geq 2$ and $\ell \geq 1$. Furthermore, $1 + \ell/k^\ell \leq R(k, \ell) \leq 2$.*

Proof. Define S to be the set of all real numbers $\alpha \geq 1$ such that there exists a (α, ℓ)-free infinite word over a k-letter alphabet. Since Thue proved that there exists an infinite word over a two-letter alphabet (and hence over larger alphabets) avoiding $(2+\epsilon)$-powers for all $\epsilon > 0$, we have that $\beta = \inf S$ exists and $\beta \leq 2$. If $\beta \in S$, we are in case (b) above, and if $\beta \notin S$, we are in case (a). Thus $R(k, \ell) = \beta$.

For the lower bound, note that any word of length $\geq k^\ell + \ell$ contains $\geq k^\ell + 1$ factors of length ℓ. Since there are only k^ℓ distinct factors of length ℓ, such a word contains at least two occurrences of some word of length ℓ, and hence is not $(1 + \frac{\ell}{k^\ell}, \ell)$-free. ■

Remarks.

1. It may be worth noting that we know no instance where case (b) of the definition of generalized repetition threshold above actually occurs, but we have not been able to rule it out.

2. Using the Lovász local lemma, Beck [1] has proved a related result: namely, for all $\epsilon > 0$, there exists an integer n' and an infinite $(1 + n/(2-\epsilon)^n, n)$-free binary word for all $n \geq n'$. Thus our work can be viewed as a first attempt at an explicit version of Beck's result (although in our case the exponent does not vary with n).

3 Conjectures

In this section we give some conjectures about $R(k,\ell)$.

Conjecture 2

$$R(k,2) = \begin{cases} (3k-2)/(3k-4), & \textit{if } k \textit{ is even;} \\ (3k-3)/(3k-5), & \textit{if } k \textit{ is odd.} \end{cases}$$

Conjecture 3 *$R(3,\ell) = 1 + \frac{1}{\ell}$ for $\ell \geq 2$.*

These conjectures are weakly supported by some numerical evidence. The following table gives the established and conjectured values of $R(k,\ell)$. Entries in **bold** have been proved; other entries, in light gray, are merely conjectured. If the entry for (k,ℓ) is α, then we have proved, using the usual tree-traversal technique discussed below, that there is no infinite (α,ℓ)-free word over a k-letter alphabet.

$R(k,\ell)$	ℓ						
k	1	2	3	4	5	6	7
2	**2**	**2**	$\frac{8}{5}$	$\frac{3}{2}$	$\frac{7}{5}$	$\frac{4}{3}$	$\frac{9}{7}$
3	**$\frac{7}{4}$**	**$\frac{3}{2}$**	$\frac{4}{3}$	$\frac{5}{4}$	$\frac{6}{5}$	$\frac{7}{6}$	$\frac{8}{7}$
4	**$\frac{7}{5}$**	$\frac{5}{4}$	$\frac{6}{5}$	$\frac{7}{6}$			
5	**$\frac{5}{4}$**	$\frac{6}{5}$	$\frac{8}{7}$				
6	**$\frac{6}{5}$**	$\frac{8}{7}$					
7	**$\frac{7}{6}$**						
8	**$\frac{8}{7}$**						
9	**$\frac{9}{8}$**						
10	**$\frac{10}{9}$**						
11	**$\frac{11}{10}$**						
12	$\frac{12}{11}$						
13	$\frac{13}{12}$						

Figure 1: Known and conjectured values of $R(k,\ell)$.

The proved results are as follows:

- $R(2,2) = 1$ follows from Thue's proof of the existence of overlap-free words over a two-letter alphabet [14, 15, 2];
- $R(2,2) = 2$ follows from Thue's proof together with the observation of Entringer, Jackson and Schatz [7];

- $R(3,1) = 7/4$ is due to Dejean [5];
- $R(4,1) = 7/5$ is due to Pansiot [10];
- $R(k,1) = k/(k-1)$ for $5 \leq k \leq 11$ is due to Moulin-Ollagnier [9];
- $R(3,2) = 3/2$ is new and is proved in Section 4.

We now explain how the conjectured results were obtained. We used the usual tree-traversal technique, as follows: suppose we want to determine if there are only finitely many words over the alphabet Σ that avoid a certain set of words S. We construct a certain tree T and traverse it using breadth-first search. The tree T is defined as follows: the root is labeled ϵ (the empty word). If a node w has a factor contained in S, then it is a leaf. Otherwise, it has children labeled wa for all $a \in \Sigma$. It is easy to see that T is finite if and only if there are finitely many words avoiding S.

We can take advantage of various symmetries in S. For example, if S is closed under renaming of the letters (as is the case in the examples we study), we can label the root with an arbitrary single letter (instead of ϵ) and deduce the number of leaves in the full tree by multiplying by $|\Sigma|$.

If the tree is finite, then certain parameters about the tree give useful information about the set of finite words avoiding S:

- If h is the height of the tree, then any word of length $\geq h$ over Σ contains a factor in S.
- If M is the length of longest word avoiding S, then $M = h - 1$.
- If I is the number of internal nodes, then there are exactly I finite words avoiding S. Furthermore, if L is the number of leaves, then (as usual), $L = 1 + (|\Sigma| - 1)I$.
- If I' is the number of internal nodes at depth $h - 1$, then there are I' words of maximum length avoiding S.

Table 2 gives the value of some of these parameters. Here α is the established or conjectured value of $R(k,\ell)$ from Table 1.

We have seen how to prove computationally that only finitely many (α,ℓ)-free words exist. But what is the evidence that suggests we have determined the smallest possible α? For this, we explore the tree corresponding to avoiding (α^+,ℓ)-repetitions using *depth-first* (and not breadth-first) search. If we are able to construct a "very long" word avoiding (α^+,ℓ)-repetitions, then we suspect we have found the optimal value of α. For each unproven α given in Table 1, we were able to construct a word of length at least 500 (and in some cases, 1000) avoiding the corresponding repetitions. This constitutes weak evidence of the correctness of our conjectures, but it is evidently not conclusive.

To show what can go wrong, the data we presented evidently suggests the conjecture $R(2,\ell) = (\ell+2)/\ell$. But we have proven this is not true, since the tree avoiding $(1.2608, 8)$-repetitions is finite, with height 195 and 53699993 internal nodes. (Perhaps $R(2,8) = 29/23$.)

k	ℓ	α	L	I	h	$M = h-1$	I'	lexicographically least word of length M avoiding (α, l)-repetitions
2	1	**2**	8	7	4	3	2	010
2	2	**2**	478	477	19	18	2	010011000111001101
2	3	8/5	5196	5195	34	33	12	001100001010111100001110101000110
2	4	3/2	13680	13679	54	53	4	01110010010111110000011011010010011111000001011011 0001
2	5	7/5	40642	40641	60	59	4	00111010101000001111110010001011101100000011111010101000001
2	6	4/3	21476	21475	40	39	4	000110101101000000001111111010101001000110
2	7	9/7	81368	81367	65	64	4	0001111011100000000101010101111111100100100101101101100000000 1010
3	1	**7/4**	6393	3196	39	38	18	01020121021201021012021020121021201020
3	2	**3/2**	11655	5827	31	30	6	012002112201100221120011022012
3	3	4/3	4037361	2018680	228	227	6	0121210001112220101212000222101020211122200012120201110002121010222000111202010122211102021210001112220101212000222111202010122200011102021210002220101212000111222101020211100012120201112220002121010222111202010122200011102020 1
3	4	5/4	188247	94123	63	62	24	00102202111100001221210200201111222210010120220211100001212210
3	5	6/5	493653	246826	63	62	12	01011121200000222221110102020212121000001111122022002101210120
3	6	7/6	782931	391465	60	59	24	0000121121210202022011111100000021222121201010110 22222200001
3	7	8/7	2881125	1440562	68	67	24	0000111111122202020101010121212120000000022222221101101001202021 2021
4	1	**7/5**	709036	236345	122	121	48	0120310213012310320130210312301320310213012032102312013021032012310213203123013210231203213012310320130210312301320310230
4	2	5/4	10324	3441	17	16	24	0112330022110332
4	3	6/5	153724	51241	24	23	96	01012333000222111332001
4	4	7/6	2501620	833873	35	34	24	0010122223033111100002212333301011
4	5	8/7	30669148	10223049	40	39	864	001012222230331111100000221233333010101
5	1	**5/4**	1785	446	7	6	120	012340
5	2	6/5	453965	113491	23	22	240	0122344002114332204413
5	3	8/7	7497345	1874336	34	33	720	010123234440002111433322204041312
6	1	**6/5**	13386	2677	8	7	720	0123450
6	2	8/7	3159066	631813	21	20	1440	01233455002211443052
7	1	**7/6**	112441	18740	9	8	5040	01234560
8	1	**8/7**	1049448	149921	10	9	40320	012345670

Figure 2: Tree statistics for various values of k and l

4 A new result

In this section we prove the following new result:

Theorem 4 $R(3,2) = \frac{3}{2}$.

Proof. From the numerical results reported in Table 2, we know that there exist no infinite words over a 3-letter alphabet avoiding $(\frac{3}{2}, 2)$-repetitions. It therefore suffices to exhibit an infinite word over a 3-letter alphabet that avoids $(\frac{3}{2}^+, 2)$-repetitions.

Now consider the uniform morphism $h : \{0,1,2,3\}^\star \longrightarrow \{0,1,2\}^\star$ defined by

$$\begin{aligned} h(0) &= 000211, \qquad & h(2) &= 020011, \\ h(1) &= 101221, \qquad & h(3) &= 120221. \end{aligned}$$

By a result of [10], there exist $\frac{7}{5}^+$-free infinite words over four letters. Consider one such word $\mathbf{x}$. We will prove that $h(\mathbf{x})$ is $(\frac{3}{2}, 2)^+$-free.

We notice first the following synchronising property of the morphism h: for any $a, b, c \in \{0,1,2,3\}$ and $s, r \in \{0,1,2\}^\star$, if $h(ab) = rh(c)s$, then either $r = \varepsilon$ and $a = c$ or $s = \varepsilon$ and $b = c$. This is straightforward to verify.

We now argue by contradiction. Assume $h(\mathbf{x})$ has a repetition xyx such that $|x| > |y|$. If $|x| \geq 11$, then each occurrence of x contains as factor at least one full h-image of a letter. By the above synchronising property, the second x will contain the same full images and at the same positions, say $x = x'x''x'''$ with $x'' = h(u)$, $|x'| \leq 5, |x'''| \leq 5$. Therefore, $h(\mathbf{x})$ contains the factor $x'h(u)x'''yx'h(u)x'''$ and $\mathbf{x}$ has the factor uvu, where $h(v) = x'''yx'$. We next compute the order of this repetition in $\mathbf{x}$. Assuming $|x| \geq 50$, we have $|x| \geq 5|x'x'''|$ and so

$$\frac{|uvu|}{|uv|} = 1 + \frac{|x''|}{|x| + |y|} = 1 + \frac{|x| - |x'x'''|}{|x| + |y|} > \frac{7}{5},$$

a contradiction, since $\mathbf{x}$ is $\frac{7}{5}^+$-free. For the case when $|x| < 50$, it can be shown by exhaustive search that the only possibility for such a repetition in $h(\mathbf{x})$ is $|xy| = 1$. Thus, $h(\mathbf{x})$ is $(\frac{3}{2}, 2)^+$-free. ■

Remarks.

1. We needed much less than $|x| > |y|$ in obtaining the contradiction. In fact, $\frac{|x|}{|y|} > \frac{2}{3} + \delta$, for some $\delta > 0$ is sufficient. What we obtain is, for any $\delta > 0$ there exists $k = k(\delta)$ such that $h(\mathbf{x})$ is $(\frac{7}{5} + \delta, k(\delta))$-free.

2. The set of those x, y for which we need to check $(\frac{3}{2}, 2)^+$-freeness can be substantially reduced by a more detailed analysis. (For instance, we bounded $|x'x'''|$ by 10, the simplest bound, but this can be significantly reduced.) We used $|x| < 50$ in order to simplify the proof.

5 Acknowledgments

We thank Narad Rampersad for a careful reading of this paper and for pointing out the remark of Pansiot.

References

[1] J. Beck. An application of Lovász local lemma: there exists an infinite 01-sequence containing no near identical intervals. In A. Hajnal, L. Lovász, and V. T. Sós, editors, *Finite and Finite Sets*, Vol. 37 of *Colloq. Math. Soc. János Bolyai*, pp. 103–107. 1981.

[2] J. Berstel. *Axel Thue's Papers on Repetitions in Words: a Translation.* Number 20 in Publications du Laboratoire de Combinatoire et d'Informatique Mathématique. Université du Québec à Montréal, February 1995.

[3] F.-J. Brandenburg. Uniformly growing k-th power-free homomorphisms. *Theoret. Comput. Sci.* **23** (1983), 69–82.

[4] C. Choffrut and J. Karhumäki. Combinatorics of words. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, Vol. 1, pp. 329–438. Springer-Verlag, 1997.

[5] F. Dejean. Sur un théorème de Thue. *J. Combin. Theory. Ser. A* **13** (1972), 90–99.

[6] F. M. Dekking. On repetitions of blocks in binary sequences. *J. Combin. Theory. Ser. A* **20** (1976), 292–299.

[7] R. C. Entringer, D. E. Jackson, and J. A. Schatz. On nonrepetitive sequences. *J. Combin. Theory. Ser. A* **16** (1974), 159–164.

[8] A. S. Fraenkel and R. J. Simpson. How many squares must a binary sequence contain? *Electronic J. Combinatorics* **2** (1995), #R2.

[9] J. Moulin-Ollagnier. Proof of Dejean's conjecture for alphabets with $5, 6, 7, 8, 9, 10$ and 11 letters. *Theoret. Comput. Sci.* **95** (1992), 187–205.

[10] J.-J. Pansiot. A propos d'une conjecture de F. Dejean sur les répétitions dans les mots. *Disc. Appl. Math.* **7** (1984), 297–311.

[11] H. Prodinger and F. J. Urbanek. Infinite 0–1-sequences without long adjacent identical blocks. *Discrete Math.* **28** (1979), 277–289.

[12] N. Rampersad, J. Shallit, and M.-w. Wang. Avoiding large squares in infinite binary words. In *Proceedings of Words '03: 4th International Conference on Combinatorics on Words*, pp. 185–197. 2003. Turku Centre for Computer Science, TUCS General Publication #27. Paper available at `http://www.arxiv.org/abs/math.CO/0306081`.

[13] J. Shallit. Simultaneous avoidance of large squares and fractional powers in infinite binary words. To appear, *Int. J. Found. Comput. Sci.*. Preprint available at `http://www.arxiv.org/abs/math.CO/0304476`, 2003.

[14] A. Thue. Über unendliche Zeichenreihen. *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **7** (1906), 1–22. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 139–158.

[15] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **1** (1912), 1–67. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 413–478.